Multi-path in Structured Light Scanning

Daniel L. Lau*

Abstract

This document describes how the multi-path problem affects phase-based structured light systems.

1. Structured Light

Three-dimensional surface scanning by means of structured light is performed using a series of striped patterns projected onto a target scene and captured by a digital camera, placed at a triangulation angle of the projector's line of sight. The pixels of the captured images are then processed to identify a unique projector row coordinate for which the subject camera pixel corresponds. Perhaps one of the simplest means of SLI is through the use of phase-shift keying where the component patterns are defined by the set, $\{I_n^p : n = 0, 1, \ldots, N-1\}$, according to:

$$I_n^p(x^p, y^p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi(\frac{n}{N} - y^p)\right). \tag{1}$$

where (x^p, y^p) is the column and row coordinate of a pixel in the projector, I_n^p is the intensity of that pixel in a projector with dynamic range from 0 to 1, and n represents the phase-shift index over the N total patterns.

For reconstruction, a camera captures each image where the sine wave pattern is distorted by the scanned surface topology, resulting in the patterned images expressed as:

$$I_n^c(x^c, y^c) = A^c + B^c \cos\left(\frac{2\pi n}{N} - \theta\right). \tag{2}$$

where (x^c, y^c) is the coordinates of a pixel in the camera while $I_n^c(x^c, y^c)$ is the intensity of that pixel. The term A^c is the averaged pixel intensity across the pattern set that includes the ambient light component, which can be derived according to:

$$A^c = \frac{1}{N}\sum_{n=0}^{N-1} I_n^c(x^c, y^c). \tag{3}$$

Correspondingly, the term B^c is the intensity modulation of a given pixel and is derived from $I_n^c(x^c, y^c)$ in terms of real and imaginary components where:

$$B_{\mathcal{R}}^c = \sum_{n=0}^{N-1} I_n^c(x^c, y^c)\cos\left(\frac{2\pi n}{N}\right) \tag{4}$$

and

$$B_{\mathcal{I}}^c = \sum_{n=0}^{N-1} I_n^c(x^c, y^c)\sin\left(\frac{2\pi n}{N}\right) \tag{5}$$

such that

$$B^c = \|B_{\mathcal{R}}^c + jB_{\mathcal{I}}^c\| = \left\{{B_{\mathcal{R}}^c}^2 + {B_{\mathcal{I}}^c}^2\right\}^{\frac{1}{2}}, \tag{6}$$

which is the amplitude of the observed sinusoid.

If $I_n^c(x^c, y^c)$ is constant or less affected by the projected sinusoid patterns, B^c will be close to zero. Thus B^c is employed as a shadow noise detector/filter [?] such that the shadow-noised regions, with small B^c values, are discarded from further processing. Of the reliable pixels with sufficiently large B^c, θ represents the phase value of the captured sinusoid pattern derived as:

$$\theta = \angle(B_{\mathcal{R}}^c + jB_{\mathcal{I}}^c) = \arctan\left\{\frac{B_{\mathcal{I}}^c}{B_{\mathcal{R}}^c}\right\}, \tag{7}$$

which is used to derive the projector row according to $\theta = 2\pi y^p$.

2. Multi-Path

In signal processing, it is often times convenient to assume a sample of an analogue signal is its value at an infinitesimally thin sliver of time, but in fact, a sample is the average value of the signal over a fixed interval in time. In digital cameras, a pixel collects light over a fixed angle in the horizontal,

θ, and vertical, ϕ. As such, an accurate model of a pixel is not eqns. (4) and (5) but by:

$$B^c_{\mathcal{R}} = \sum_{n=0}^{N-1} \int_\theta \int_\phi I^c_n(\theta, \phi) \cos\left(\frac{2\pi n}{N}\right) d\theta d\phi \tag{8}$$

and

$$B^c_{\mathcal{I}} = \sum_{n=0}^{N-1} \int_\theta \int_\phi I^c_n(\theta, \phi) \sin\left(\frac{2\pi n}{N}\right) d\theta d\phi. \tag{9}$$

In this form, we can now identify the principal problem of multi-path, which occurs when $I^c_n(\theta, \phi)$ corresponds to a foreground object for same range on θ and ϕ and a background object for the rest of the θ and ϕ within the field of view of the subject pixel. We can describe this mathematically according to:

$$B^c_{\mathcal{R}} = \sum_{n=0}^{N-1} \int_{\theta_f} \int_{\phi_f} I^c_n(\theta, \phi) \cos\left(\frac{2\pi n}{N}\right) d\theta_f d\phi_f + \sum_{n=0}^{N-1} \int_{\theta_b} \int_{\phi_b} I^c_n(\theta, \phi) \cos\left(\frac{2\pi n}{N}\right) d\theta_b d\phi_b \tag{10}$$

and

$$B^c_{\mathcal{I}} = \sum_{n=0}^{N-1} \int_{\theta_f} \int_{\phi_f} I^c_n(\theta, \phi) \sin\left(\frac{2\pi n}{N}\right) d\theta_f d\phi_f + \sum_{n=0}^{N-1} \int_{\theta_b} \int_{\phi_b} I^c_n(\theta, \phi) \sin\left(\frac{2\pi n}{N}\right) d\theta_b d\phi_b \tag{11}$$

where θ_f and ϕ_f represent the range of θ and ϕ covering the foreground object while θ_b and ϕ_b cover the background object. We can simplify both these equations by writing:

$$B^c_{\mathcal{R}} = B^{c,f}_{\mathcal{R}} + B^{c,b}_{\mathcal{R}} \tag{12}$$

and

$$B^c_{\mathcal{I}} = B^{c,f}_{\mathcal{I}} + B^{c,b}_{\mathcal{I}} \tag{13}$$

where we added the superscripts f and b to distinguish between the foreground and background components on $B^c_{\mathcal{R}}$ and $B^c_{\mathcal{I}}$.

Now notice that, many times, scanners use multiple spatial frequencies to encode depth such that they can minimize the effect of thermal noise in

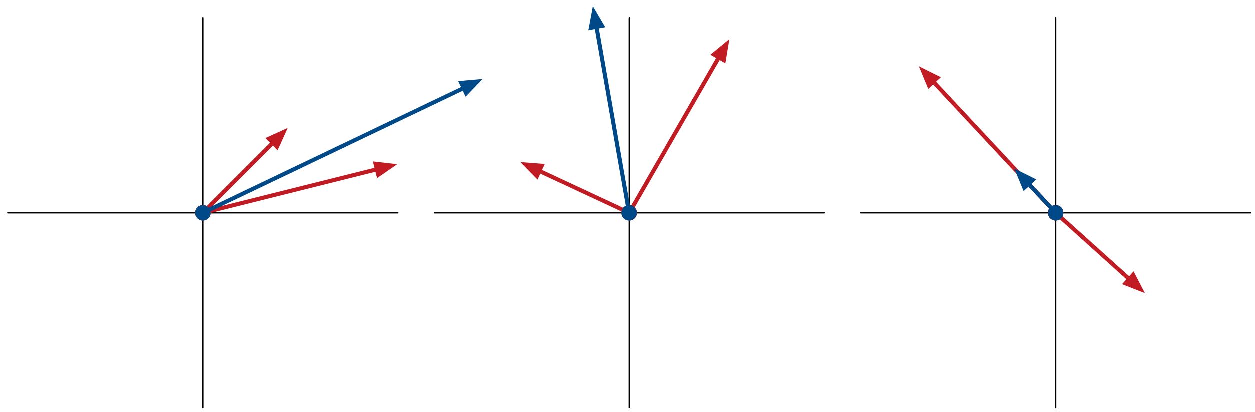

Fig. 1. Illustration of the change in direction and magnitude in the (blue) observed complex vector $B^c_{\mathcal{R}} + jB^c_{\mathcal{I}}$ created by the superposition of (red) complex vectors from multi-path fore and background objects for $K = 1$, $K = 8$, and $K = 12$.

the camera. Mathematically, increasing the spatial frequency of the PMP patterns by a factor of K increased the phase term by an equal amount while keeping the amplitude of the sinusoid constant. In the case of multi-path, this frequency scaling has a far different affect as illustrated graphically in Fig. 1 where we show (left) the fore and background components assuming unit frequency while (center) and (right) show the same components when $K = 8$ and 12.

What Fig. 1 (left) shows in red are the complex vectors formed by $B^{c,f}_{\mathcal{R}}$ and $B^{c,f}_{\mathcal{I}}$ and $B^{c,b}_{\mathcal{R}}$ and $B^{c,b}_{\mathcal{I}}$, while the blue vector shows the superimposed vectors forming the single vector formed by $B^c_{\mathcal{R}}$ and $B^c_{\mathcal{I}}$. Now note that by using a frequency scaling of K, we expect the direction or phase of the foreground and background vectors to scale by an equal amount. Graphically, this is depicted by a rotation of the vectors around the origin. Notice, though, that by rotating the vectors separately, that it is quite likely that the phase of the combined vectors are not equal to the scaling of the phase term prior to frequency scaling. Likewise, the vectors may swing from constructively interfering where magnitude of the combine vectors is equal to the sum of the individual magnitudes to destructively interfering where the magnitude of the combine vectors is equal to the difference of the individual magnitudes.

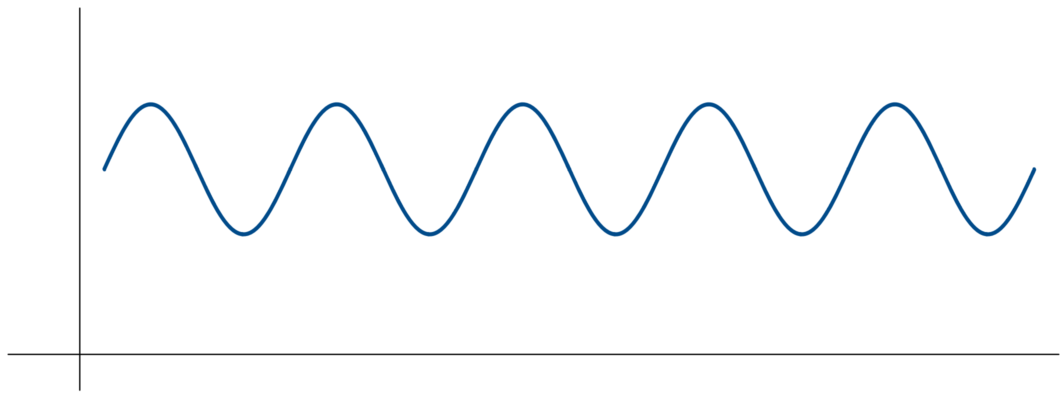

Fig. 2. Illustration of the change in magnitude in the observed complex vector $B^c_{\mathcal{R}} + jB^c_{\mathcal{I}}$ as a function of the scaling factor K.

The change in vector phase and magnitude in the superimposed vectors as a function of K is the prime means by which we intend to detect multi-path in the scanned image. We, furthermore, intend to extract the component vectors by looking at the change in phase and magnitude in the combined vectors as we vary K over multiple scans. We depict this in Fig. 2 where we plot the magnitude in the combined vector as a function of K where we can clearly see that the mean value in magnitude over all K is equal to the magnitude of the larger component (foreground or background). The difference in height between the peaks and valleys gives us the magnitude of the smaller vector. We can, likewise, use the position of the peaks and valleys to determine how much faster/slower the smaller vector is rotating around the larger vector, telling us the difference in phase between the two, which is enough to separate the two vectors and, thereby, make two separate measurements.

3. Multi-Texture

The multi-texture is very similar to the multi-path problem. Here, a single pixel sees a continuous smooth surface, but the surface texture has a discontinuity or step edge mid-way across the pixel's field of view. We can define the brighter side of the edge as the foreground surface while the darker side of the edge as the background surface. This means that the phase values inside the foreground surface will have a greater weight, per unit area, than the background surface. And this has the effect of pushing the combined vector closer to the foreground phase than the background. While the change may not be as severe as the multi-path problem, the solution is the same, look at the change in phase and magnitude of the measured vector as a function of K.